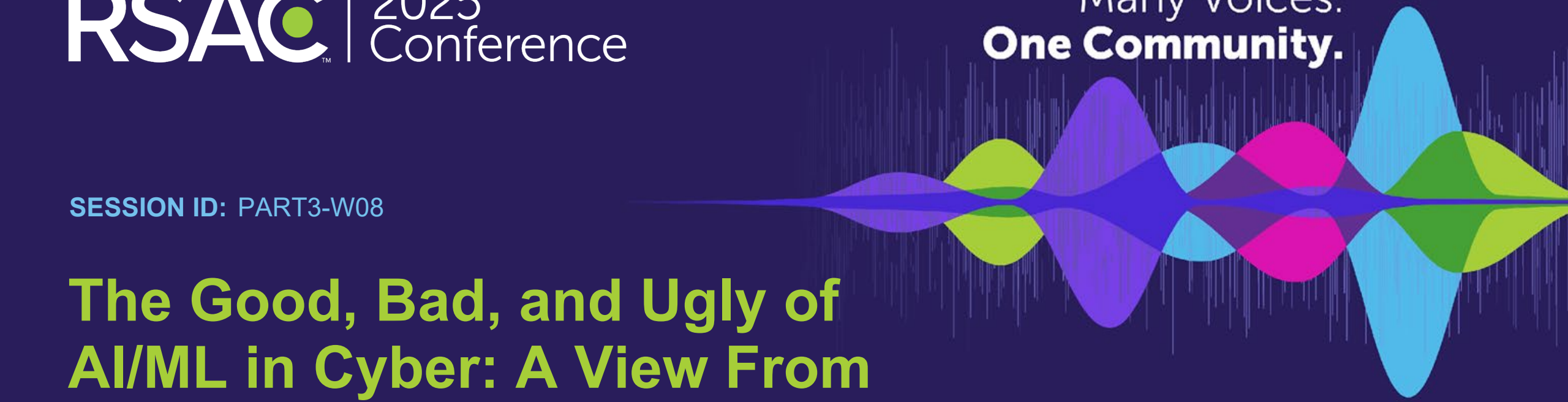
RSAC 2025 Conference

Many Voices. One Community.



SESSION ID: PART3-W08

The Good, Bad, and Ugly of AI/ML in Cyber: A View From the Front Line

Chad E. LeMaire
Deputy CISO
ExtraHop
https://www.linkedin.com/in/chad-lemaire-ciso/

Bob Skinner
President
Skinner's Strategic Solutions, LLC
https://www.linkedin.com/in/robert-skinner-1a48aa22/

Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.

Agenda

- Introduction
- The Ugly
- The Bad
- The Good
- Perspectives
- Questions

Introduction

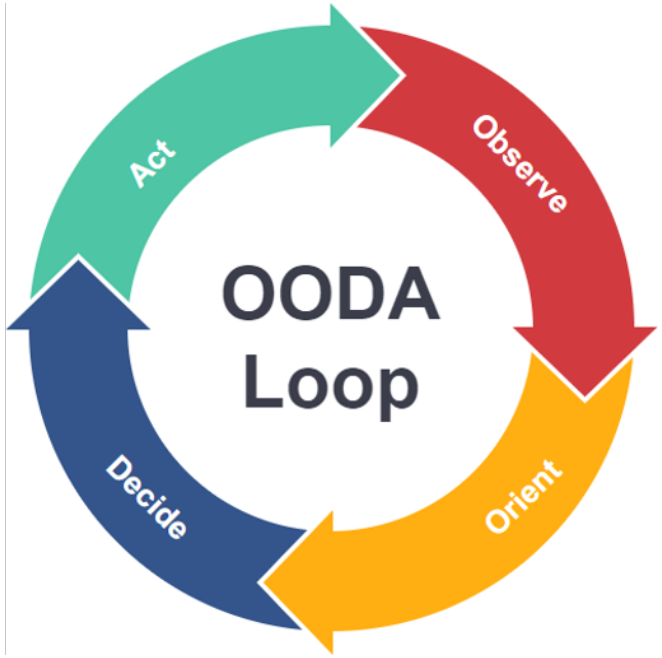
- Our Backgrounds
- Asymmetric Warfare
- OODA Loop
- Changing Landscape

Asymmetric Warfare

- Velocity vs Speed
- Disparity in sophistication
- Disparity in costs
- Unconventional solutions

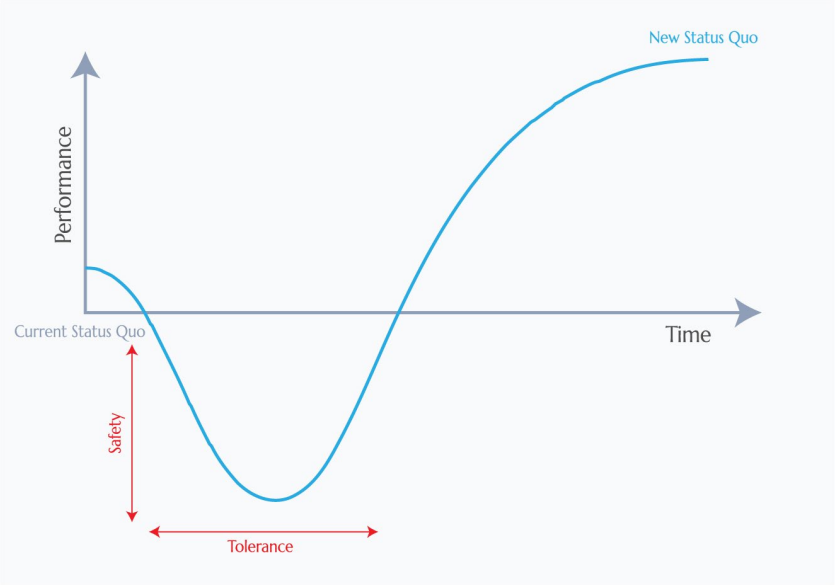
OODA Loop

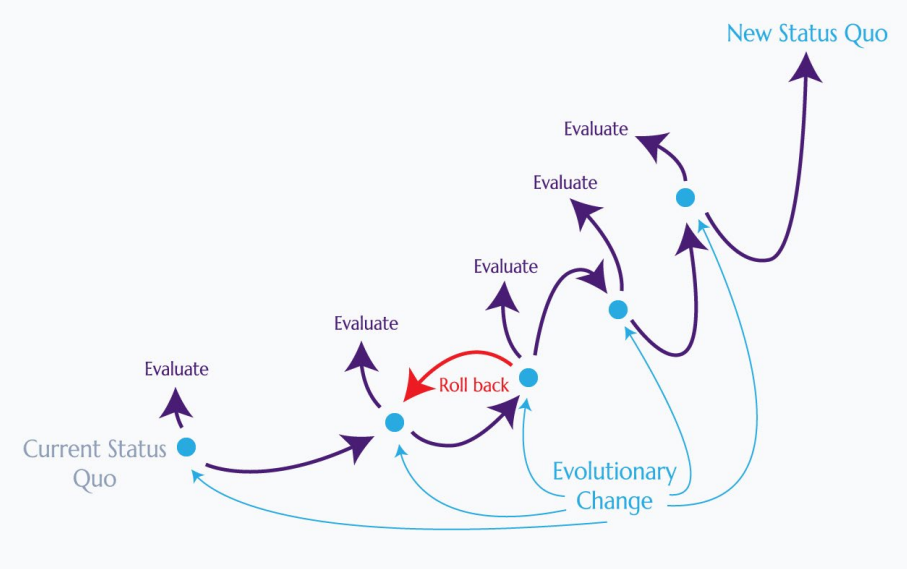
- Concept developed in the 1950s, published in 1986
  - Proven concept
- Observe, Orient, Decide, Act
- AI is greatly speeding up Orient, Decide, and Act
  - Race between threat actors and defenders



Changing Landscape

- Quality or Quantity vs Quality and Quantity
- J-curve and Exponential Growth
- Scope and Scale





Images from: https://getnave.com/blog/j-curve-effect/#What-is-the-J-Curve-Effect

RSAC | 2025 Conference

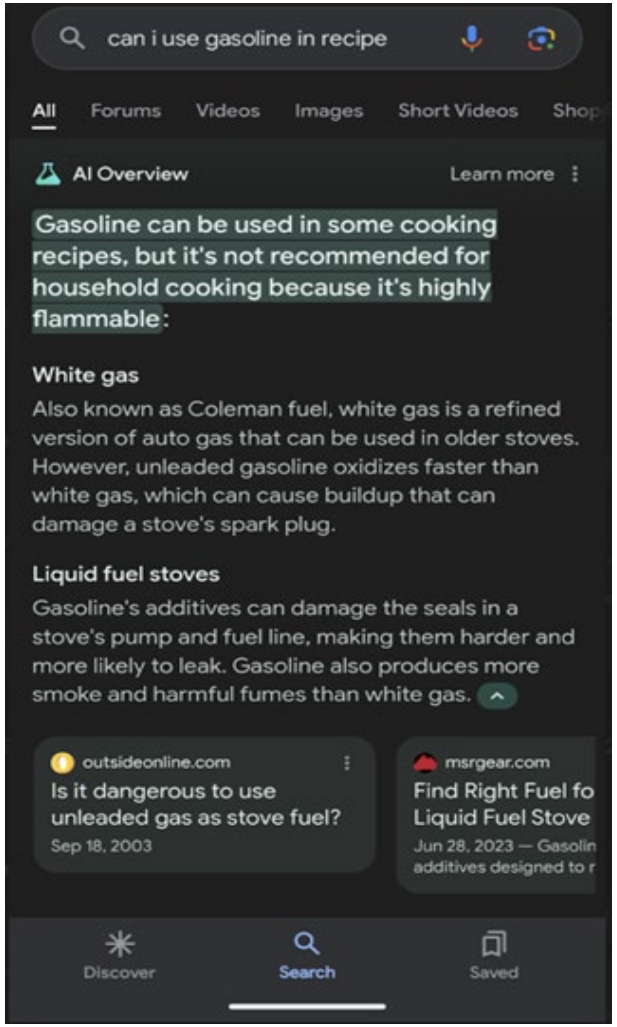
The Ugly of AI/ML in Cyber

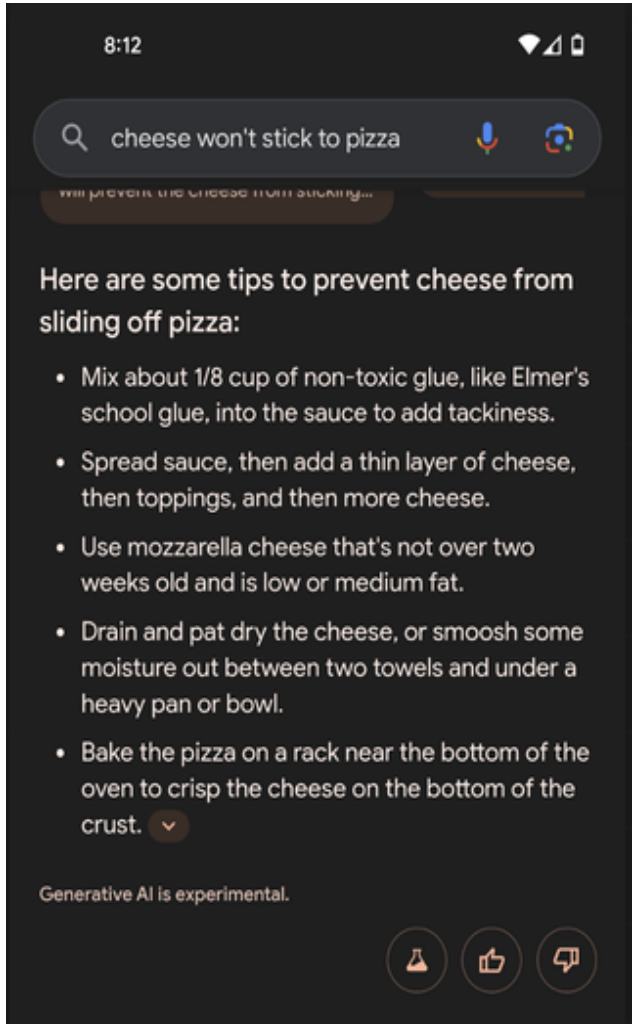
The Ugly of AI/ML

- AI Hallucination
- Model/Data Drift
- Model/Data Poisoning
- AI Adoption and Understanding

The Ugly of AI/ML

how to remember your pass...
All
Videos
Images
Short Videos
Forums
AI Overview
Learn more
Here are some tips for remembering your password:
Try variations of your name: Try your name, along with other personal details like your birthday.
Use a passphrase: A passphrase is four or more random words that are as strong as a password that uses a mix of letters, numbers, and symbols.
Make it long: Longer passwords are stronger.
Use poetic language: Write down the first letter of each syllable, using capital letters for stressed syllables.
Include all four types of characters: Use uppercase letters, lowercase letters, digits, and symbols.
NordPass
Remember forgotten passwords FAST | NordPass
Jul 13, 2020
PCMag
3 Simple Tricks Remembering
Discover
Search
Saved


can i use gasoline in recipe
All
Forums
Videos
Images
Short Videos
AI Overview
Learn more
Gasoline can be used in some cooking recipes, but it's not recommended for household cooking because it's highly flammable:
White gas
Also known as Coleman fuel, white gas is a refined version of auto gas that can be used in older stoves. However, unleaded gasoline oxidizes faster than white gas, which can cause buildup that can damage a stove's spark plug.
Liquid fuel stoves
Gasoline's additives can damage the seals in a stove's pump and fuel line, making them harder and more likely to leak. Gasoline also produces more smoke and harmful fumes than white gas.
outsideonline.com
Is it dangerous to use unleaded gas as stove fuel?
Sep 18, 2003
msrgear.com
Find Right Fuel fo Liquid Fuel Stove
Discover
Search
Saved


8:12
cheese won't stick to pizza
Here are some tips to prevent cheese from sliding off pizza:
Mix about 1/8 cup of non-toxic glue, like Elmer's school glue, into the sauce to add tackiness.
Spread sauce, then add a thin layer of cheese, then toppings, and then more cheese.
Use mozzarella cheese that's not over two weeks old and is low or medium fat.
Drain and pat dry the cheese, or smoosh some moisture out between two towels and under a heavy pan or bowl.
Bake the pizza on a rack near the bottom of the oven to crisp the cheese on the bottom of the crust.
Generative AI is experimental.

RSAC | 2025 Conference

The Bad of AI/ML in Cyber

The Bad of AI/ML

- AI/ML meets the DarkWeb:
  - Script kiddie for script kiddies?
- Threat actor poisoning of AI models
- Sophistication
  - Social Engineering is becoming harder to distinguish
  - Implants
- Commoditization of malware / low cost of entry for threat actors
  - $5K-$20K entry; 2024 average ransom payment was $479K!

The Bad of AI/ML - Continued

- Cost of complex defense systems can be > $500K
- Potential to expose intellectual property
- Deep access: dormant until a future event / time
  - Threat actor risk is lower
- Next level is AI creating AI: AI agents / “Agentful”

RSAC | 2025 Conference

The Good of AI/ML in Cyber

The Good of AI/ML

- AI aware defense capabilities
- ML: ability to analyze millions of data points
  - Situational Awareness: understanding your environment
  - Anticipating changes in the environment and threat landscape
- Scale
- Unconventional solutions

RSAC 2025 Conference

Perspectives

Perspectives: The Basics Are Changing

- ”Basic” Blocking and Tackling is Foundational
- Sophistication of attacks means the basics must be right
  - Have we truly seen the power of AI assisting threat actors?
- Will Moderate and Advanced defenses become the basics of tomorrow?

Perspectives: Identity is the Center of Gravity

- Identity is the key attack surface
- Trust is king; walk (or run) through the front door
- Two-factor is not enough in the AI environment

Perspectives: OODA Loop Must Shrink

- Veracode reports:
  - 252 days to mitigate security flaws: 50% more than 1 year
  - Most are from 3rd party and software supply chain
- AI/ML will impact each OODA Loop Phase
- Threat actors will have initial advantage

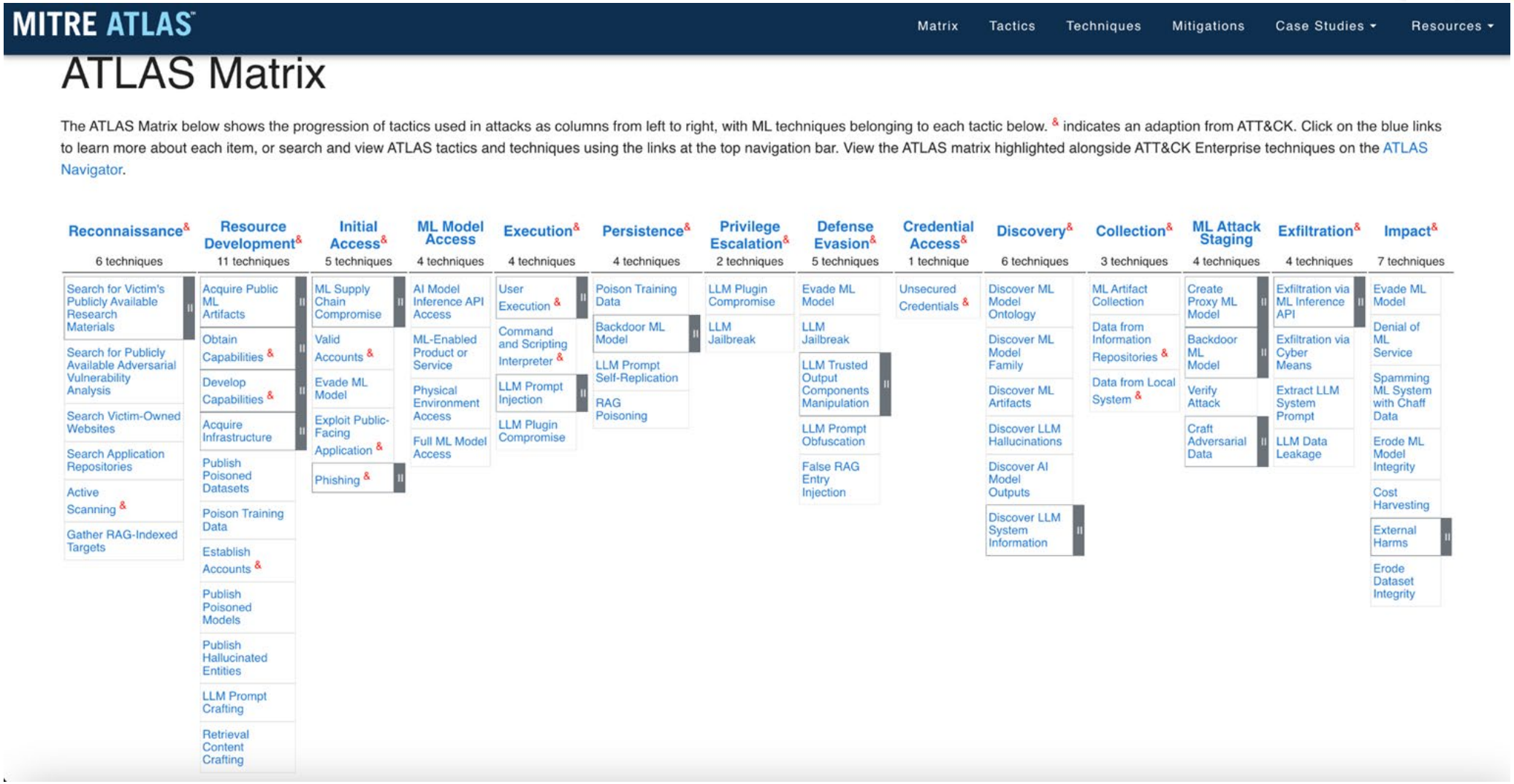
Perspectives: Automation and AI Are Not Necessarily the Same Thing

- Automation is Centuries old
  - Automatic Flour Mill (1785)
  - Computer Processing Automation (1990)
- Artificial Intelligence is Decades Old
  - Logic Theorist (1955)
  - Really Popular in 2023 … mainstream
- Automation is Rudimentary … sub-optimized without process change

Perspectives: It Takes AI to Defeat AI

- Time
- Scale
- Complexity

Perspectives: MITRE ATLAS



MITRE ATLAS

Matrix Tactics Techniques Mitigations Case Studies Resources

ATLAS Matrix

The ATLAS Matrix below shows the progression of tactics used in attacks as columns from left to right, with ML techniques belonging to each tactic below. & indicates an adaption from ATT&CK. Click on the blue links to learn more about each item, or search and view ATLAS tactics and techniques using the links at the top navigation bar. View the ATLAS matrix highlighted alongside ATT&CK Enterprise techniques on the ATLAS Navigator.

Reconnaissance&	Resource Development&	Initial Access&	ML Model Access	Execution&	Persistence&	Privilege Escalation&	Defense Evasion&	Credential Access&	Discovery&	Collection&	ML Attack Staging	Exfiltration&	Impact&
6 techniques	11 techniques	5 techniques	4 techniques	4 techniques	4 techniques	2 techniques	5 techniques	1 technique	6 techniques	3 techniques	4 techniques	4 techniques	7 techniques
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	AI Model Inference API Access	User Execution &	Poison Training Data	LLM Plugin Compromise	Evade ML Model	Unsecured Credentials &	Discover ML Model Ontology	ML Artifact Collection	Create Proxy ML Model	Exfiltration via ML Inference API	Evade ML Model
Search for Publicly Available Adversarial Vulnerability Analysis	Obtain Capabilities &	Valid Accounts &	ML-Enabled Product or Service	Command and Scripting Interpreter &	Backdoor ML Model	LLM Jailbreak	LLM Jailbreak		Discover ML Model Family	Data from Information Repositories &	Backdoor ML Model	Exfiltration via Cyber Means	Denial of ML Service
Search Victim-Owned Websites	Develop Capabilities &	Evade ML Model	Physical Environment Access	LLM Prompt Injection	LLM Prompt Self-Replication		LLM Trusted Output Components Manipulation		Discover ML Artifacts	Data from Local System &	Verify Attack	Extract LLM System Prompt	Spamming ML System with Chaff Data
Search Application Repositories	Acquire Infrastructure	Exploit Public-Facing Application &	Full ML Model Access	LLM Plugin Compromise	RAG Poisoning		LLM Prompt Obfuscation		Discover LLM Hallucinations		Craft Adversarial Data	LLM Data Leakage	Erode ML Model Integrity
Active Scanning &	Publish Poisoned Datasets	Phishing &					False RAG Entry Injection		Discover AI Model Outputs				Cost Harvesting
Gather RAG-Indexed Targets	Poison Training Data								Discover LLM System Information				External Harms
	Establish Accounts &												Erode Dataset Integrity
	Publish Poisoned Models												
	Publish Hallucinated Entities												
	LLM Prompt Crafting												
	Retrieval Content Crafting												

Perspectives: NDR is Needed, not Desired

- AI threats traverse the network; network visibility is a must
- NDR + EDR + SIEM
  - Some threats evade EDR and SIEM
  - NDR sees all threats on the network
- Without NDR security teams will be overwhelmed

Summary: OODA Loop, Asymmetric Warfare, and AI

- AI will contribute to the OODA Loop as well as Asymmetric Warfare
  - As we've seen in other areas, AI will find asymmetric solutions
  - AI will enable OODA Loop completions within milliseconds
- Good versus Evil (AI is not inherently good or evil)
- The AI that combines asymmetric warfare and within a quicker OODA Loop will prevail

Application

- This week and next week you should:
  - Increase your knowledge of NDR: Security operations must include NDR in addition to EDR and SIEM
- In the next 3 months you should:
  - Study MITRE ATLAS, ATT&CK, and D3FEND knowledge bases to understand and counter threat actors and “bad” AI tactics
- You should always:
  - Protect Identity…at all costs

RSAC 2025 Conference

Many Voices.
One Community.

Questions?